Chapter 5: Importing Data Files

Data files may be imported using the SEER*DMS interface or auto-loaded from a specified network location. All data files must adhere to file specifications defined in the registry's importer configuration (refer to the *SEER*DMS Technical Reference: Registry-specific Information*).

Imported data are stored in the record table of the SEER*DMS database and can be viewed within the SEER*DMS record editor. Each record enters the workflow and initiates the automated and manual tasks which are required to process the data. You may specify the Import ID in the Patient Lookup, Worklist, or Import Manager to search for the records or to determine whether the records have completed the workflow. The Import Manager also provides summary information about each import and access to the source data files.

In this chapter, you'll learn about

- Import Manager
- Importing Data Files within SEER*DMS
- Auto-Loading Data Files
- Import Summary Information
- Reviewing Field Frequencies to Prevent Input Errors
- Import Review Task
- Identifying Records that Could Not be Loaded
- Notifying the Source Facility of Problems with Data Files
- Import Info Page
- Finding Records and Tasks Related to an Import
- Reports Related to Data Imports
- Defining New Data File Types

Import Manager

Requires system permission: *import_electronic* or *import_manual*

Use the Import Manager to locate records loaded in a particular import, to view or download a copy of the source data files, or to determine the current workflow location of each record in an import.

To access the Import Manager, select **System > Imports**. The following data columns are shown in the Import Manager:

- **Import ID** – A unique ID assigned to each import. Click the ID to view details about the import as described in the *Import Summary Information* section of this chapter.
- **Import Date** – Date and time that the import was initiated.
- **Facility** – The facility selected by the user or set in the autoloader configuration file for this import. This facility is used by default as the record's default reporting facility when the reporting facility is not available on the record.
- **User** – The user who initiated the import. "seerdms" will be listed as the user for autoloader imports and data migrated at the time of deployment.
- **Records** – The number of records imported. Thise value will be zero until the Import Review task is completed and the data are successfully loaded.
- **Type** – The import's Media Type is displayed in this column. The possible values are: Data Files, Image Files, Paper Documents, or blank (migrated data). Note: The Image Files value for Media Type will be removed in SEER*DMS version 4.0. This designation was intended for data entry from data submitted to the registry via PDF or other electronic

format. It is not used for autoloaded image files which attach scanned document files to SEER*DMS records and patient sets.

- **Comment** – The import-wide comment entered on the same page as contact when starting a new import or entered in the <> tag in the autoloader XML.
- **Action** – Use the Lookup link in the Action column to access the Patient Lookup and search the database for records loaded in the import. A maximum of 500 records will be listed. Use RPT-009A to generate a complete list of the records loaded in an import.

Importing Data Files within SEER*DMS

Requires system permission: *import_electronic* and *contact_view*

You may use the SEER*DMS interface to import data files from any local or network disk drive that you can access. The files will be uploaded through your Web browser and are subject to HTTP size limitations. Use the auto-load feature to import data files that exceed these limitations.

*To import data files using the SEER*DMS interface:*

1. Select **System > Imports** to access the Import Manager.
2. Click **New Import**.
3. Use the lookup for the **Contact (Facility)** field to specify the import facility and a contact at that facility.
4. Set the **Media Type** to *Data Files*.
5. Enter the **Reference/Barcode** if applicable for this import.
6. You may enter the date that the facility sent the data to the registry in the **Sent Date** field. This field is optional if it is not known or tracked by your registry.
7. **Received Date** is automatically set to today's date. You may override this value if the file(s) arrived at the registry on a previous date. Click the Calendar icon to modify the date, if necessary.
8. You may enter notes regarding the import in the **Comments** field.
9. Click **Next**.
10. In the **Add File** section of the wizard, click the **Browse** button. Explore to the appropriate folder and click the data file that you wish to import.
11. Select the appropriate **File Type**. You may use the lookup to review the record layout of each file type.
12. You may enter a **Comment** regarding this specific file.
13. Click **Add**. SEER*DMS will determine if the file was loaded before. If the file is a duplicate, an error message will be displayed. To remove the file from the import, click the **Remove** link in the **Action** column. (SEER*DMS will allow you to over-ride the error and load a duplicate file. Use this feature cautiously and only after consultation with registry managers and technical support staff).
14. Repeat steps 10-13 for each file which is being imported from this facility. Each file will be listed in the **Current Files** section as it is added.
15. Click **Import**.

16. SEER*DMS will analyze the data files and create a manual Import Review task prior to storing the data in the database. Refer to the *Import Review Task* section of this chapter for instructions on completing this task. The Import Review Task provides an error summary and frequencies of all data fields. Once you review these statistics, you will have the option of loading the data or terminating the import for some or all data files.

Auto-Loading Data Files

SEER*DMS can be configured to automatically load data files stored in designated locations on the network. On a regular basis, an automatic workflow task imports all files in the auto-loader directories. The frequency of the directory scan is controlled by registry-defined configuration parameters. The auto-loader allows the registry to accept routine transmissions from sources such as ePath without human intervention; and provides a mechanism for loading files that exceed HTTP size limitations. Please refer to the *SEER*DMS Technical Reference* for more information in configuring the auto-loader.

SEER*DMS will analyze the data file and create a manual Import Review task prior to loading the data into the database (refer to the *Import Review Task* section of this chapter for detailed instructions). The task provides an error summary and a listing showing the frequency of each value for all data fields. Once you review these statistics, you will have the option of loading the data or terminating the import for some or all data files.

SEER*DMS checks imported and auto-loaded data files to determine if the files were previously loaded. If you attempt to import a file that was previously loaded, SEER*DMS will display an error message and discontinue the import of that file. If a duplicate file is copied to an auto-load directory, the file will not be loaded and email notification will be sent to the user specified in the auto-loader configuration.

Import Summary Information

The following information is listed for each file on the Import Info page and in Import Review tasks.

- **Filename** – In the Import Info page, click the filename to open or download the file. In an Import Review task, the filename will only be a link if there are errors in the file. Click the link to view the Import Problems popup window.
- **Rows** – The total number of records in the data file.
- **Errors** – The number of records which cannot be evaluated due to errors in the record layout, fields which define the record type, or key fields required when the data are loaded into the record table. SEER*DMS performs a limited amount of error checking at this point. This error count will not identify invalid values in all data fields. You must review the frequencies shown at the bottom of the page to verify data values.
- **Duplicate** – If two records within the same file are exact duplicates, one will be loaded and the other will be counted here. This count also includes records which cannot be loaded because they are exact duplicates of records loaded in an earlier import (based on hash maps). It also includes records which are not exact matches but provide the same data as previously loaded records. Non-exact duplicates may have differences in non-essential fields but provide the same data in key fields. In most registries, the check for non-exact duplicates is only applied to death certificate records.
- **Ignored** – To reduce the number of extraneous supplemental records in the database, some data types are evaluated to determine if they provide useful data. A record that is deemed as not useful is “ignored” (it is not processed further and is not loaded into the record table). In some registries, a pre-load matching algorithm is applied to specific types of supplemental records. A record processed in this match is ignored if it does not match a patient in the database. The number of ignored records also includes records which are not

loaded because of a rule in the importer algorithm for the specific record type. For example, some supplemental record types are ignored if social security number is null (e.g., this rule applies to HCFA and CMS records in some registries). To review the pre-load matching algorithms and importer rules used in your registry, please refer to *SEER*DMS Technical Reference: Registry-specific Information.*

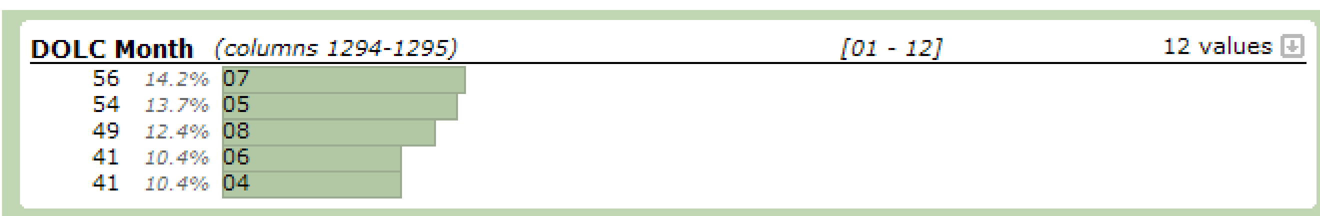
- **Valid** – The number of records which can be loaded. Valid = (Rows – (Errors + Duplicates + Ignored)). If you complete the import, this is the number of new rows that will be created in the record table and the number of records that will enter the workflow.

Reviewing Field Frequencies to Prevent Input Errors

Requires system permission: *import_resolution*

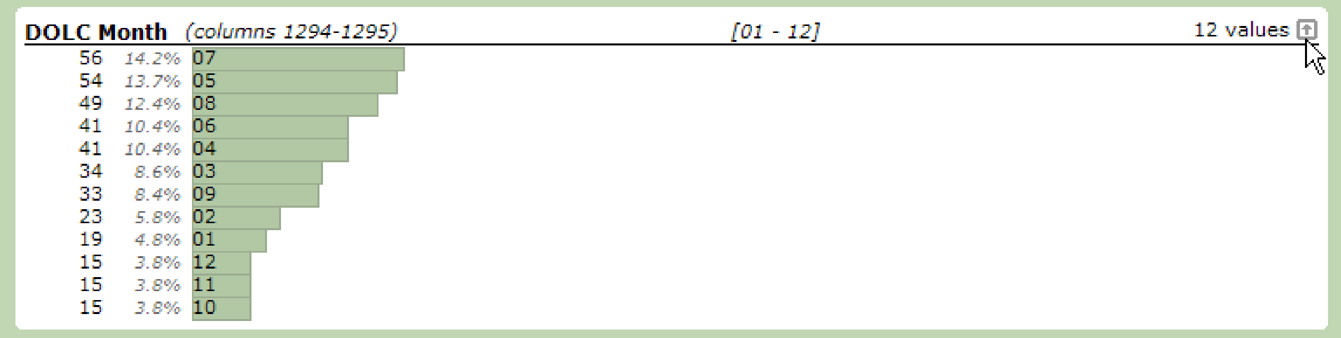
It is critical that data which are imported into the system are valid. Some import files need to be pre-processed, while others can be imported as-is. If there were problems with the data preparation, either at the sending facility or within the registry, an invalid import file could incorrectly modify a large volume of system data. This type of problem is very difficult to remedy once the data are loaded into SEER*DMS; each record immediately begins to move through the workflow and affect changes in patient set data. It is the responsibility of registry staff to define and implement comprehensive quality control procedures to prevent import errors. This should include but not be limited to the review of input field frequencies to identify invalid or unacceptable values. Field frequencies are provided in the Import Review task for each import file.

The Statistics section of an Import Review task lists frequencies each data field in an import file. If multiple files are included in the import, a drop-down list allows you to switch from one file to another. The fields are listed in alphabetical order. You may use the linked letters at the top of the list to view fields which begin with that letter or use Firefox search tools to search for a fieldname or column position.



The figure above shows the type of information shown for each data field. The following describes the information that is displayed and the text actually used in the example above.

- **Field name** - shown in bold above the line: *DOLC Month*
- **Column position –** shown in parentheses next to the field name: *(columns 1294-1295)*
- **Range of values** - shown in brackets above the line in center: *[01 – 12]*
- **Number of different values** – only shown if there are more than five different values for the field. The number of different values ("*12 values*" in the example) is displayed on the far right of the page, just above the line. Up and down arrows enable you to toggle the display to show all values or hide all but the top five values. The figure below shows all twelve values for the example.
- **Statistics and bar chart -** For each value encountered for the field: the number of records, percent of records, and value. The value is shown within a histogram bar. In the example, 56 records have a value of "07" for DOLC Month. 56 equals 14.2% of the number of valid records in the file. Note: frequencies are not shown if the number of differing values exceeds display limitations. For example, it would be impractical to show a frequency for street address which would be unique for every record.



Import Review Task

Requires system permission: *import_resolution*

Once an import is initiated, SEER*DMS will analyze the data files and create a manual Import Review task prior to loading the data into the database. The Import Review Task provides an error summary and a listing showing the frequency of each value for all data fields.

To complete the Import Review tasks:

1. Click the **Import Review** link in the **My Tasks** section of the worklist summary on the home page.
2. Click the Task ID to open the task. At this stage, the files have been checked for errors and analyzed to create frequencies of values in all data fields. The records have not been loaded into the system.
3. Carefully review number of records listed for Rows, Duplicates, Ignored, and Valid in the Import Files section of the page. Each of these fields is described in the previous section. If you have external documentation related to these files, compare the number of Rows to the expected number of records for each data file.
4. Follow steps 5 and 6 for each file with errors.
5. If an incorrect file type was specified for the file:

   a. Select the correct **File Type**.

   b. Set **Action** to *Accept.*
6. If the file type is correct but the file has errors:

   a. Click the linked filename.

   b. Review the error messages. You must determine whether you want to reject all records in the file or import the records that did not cause errors. (Note: Once an error is encountered, the first record with the error and all subsequent records will be displayed in red.)

   c. Click **Close** to close the Import File Problems window.

   d. Set the appropriate **Action** for this file.

i. If you would like to contact the reporting facility and obtain a replacement for a file with errors, select *Reject* to reject all records in the file (records in other files will not be affected).

ii. If all errors are caused by duplicate records, you may use *Re-import valid only* to accept the non-duplicate records. If there are other types of errors, only use this action if you have the technical resources to identify the records with errors. You should attempt to obtain corrected versions of these records, as described in the *Identifying Records that were Not Loaded* section of this chapter.

7. To close the Import Review task:

   a. If an Action is specified for each file, click **Import** to load the acceptable data and exit the task. If the data cannot be loaded or a new file type was selected for a file, a second Import Review task will be created and assigned to you. Otherwise, the records will begin entering the workflow. Automated and manual tasks related to the records will be created in the worklist. You may specify the Import ID in the Patient Lookup, Worklist, or Import Manager to search for the records or to determine whether the records have completed the workflow. System reports provide additional information regarding the imported data.

   b. If you would like to close the Import Review task without importing any of the files:

      i. To request new files from the source facility and retain the Import Problems task as a reminder, follow the instructions in *Notifying the Source Facility of Problems with Data Files*. If you suspect the error was in a pre-processing step, request a new file from your data processing department.

      ii. Click **Reject All**.

Identifying Records that Could Not Be Loaded

As described in *Resolving Problems with Imported Data Files*, it is possible to load partial files. This will commonly be used to ignore duplicate records within a file, but may also be used to ignore records with other errors. If a partial file is loaded, it is the responsibility of the registry's technical staff to identify and obtain corrected versions of the records that were "ignored" (not loaded).

To identify the records that were not loaded, compare a list or extract of records to the original data file submitted by the reporting facility. SEER*DMS includes a system report that provides a list of the records loaded from each data file (RPT-009A). Alternatively, you may use a similar query in an external application to create a data extract.

Provide a list of the records with errors to the reporting facility and request replacement records. If you wish the request to be tracked within SEER*DMS, you may use the Record Request feature described in *Chapter 23: Requesting Records*.

Notifying the Source Facility of Problems with Data Files

Requires system permission: *import_resolution*

While performing an Import Review task, you may use the Query feature to request new data files or report errors to the organization representative.

1. Open the task as described in the *Import Review Task* section of this chapter.

2. Click **Query**. You will be able to use any communication method, provided that the relevant contact information is defined for the representative in the Contacts list. (See *Chapter 19: The Contact List* for more information.)

3. To send an e-mail:

   a. Set **Type** to E-mail.

   b. Click **To**, **CC**, or **BCC** to change the recipients.

      i. Enter text in the **Search** filter to find a specific person.

      ii. Check **To**, **CC**, or **BCC** for each contact, as appropriate. Only one contact may be selected as the main recipient (To).

      iii. Click **Save** to exit the recipient page and continue composing the e-mail message.

   c. Enter text into the **Message** box (do not include confidential information).

   d. If you would like to attach the problematic data file, click **Attach File**. Do not attach files containing confidential data unless the file is encrypted. (SEER*DMS does not provide a mechanism for encrypting data files or e-mail messages.)

   e. Click **Send**.

4. To print a letter to be faxed or mailed:

   a. Set **Type** to either Fax or Letter.

   b. Enter the text for the body of the letter into the **Message** box.

   c. Click **Print** to create a PDF file.

   d. Review the contents of the letter. If necessary, close the Acrobat window and return to the Communication page to make revisions.

   e. Use the Acrobat controls to print the letter.

5. To contact the representative by phone:

   a. Set **Type** to *Phone*.

   b. The contact's phone number will be displayed on the screen as a reference while you make the call. The message box is not relevant to phone communications.

   c. If you are unable to reach the contact, you may want to make a note to remind yourself to try again later or use an alternate method of contact. Or, click **Cancel** to cancel all changes made in this Import Problems task. The task will remain in the worklist assigned to you.

   d. Once you have successfully completed the call, click **OK**.

Import Info Page

Requires system permission: *import_electronic* or *import_manual*

You may use the Import Info page to view or download the source data files, view a summary of open worklist tasks related to the imported records, search the database for a sample of records loaded in the import, or view .

The following tools are available in the Import Info page:

- **Import ID** – A unique ID assigned to each import. This ID can be used throughout SEER*DMS to find data loaded in a particular import and is displayed when editing a record.

- **Contact** – The contact selected by the user or set in the autoloader configuration file for this import.
- **Facility** – The facility selected by the user or set in the autoloader configuration file for this import. This facility is used by default as the record's default reporting facility when the reporting facility is not available on the record.
- **Import Date** – Date and time that the import was initiated.
- **Sent Date** – Date that the data were sent by the facility to the registry. This date may not be known for the import or it may not be tracked by your registry.
- **Received Date** – Date that the data files were received at the registry.
- **User** – The user who initiated the import. "seerdms" will be listed as the user for autoloader imports and data migrated at the time of deployment.
- **Records** – The number of records imported. This value will be zero during initial pre-load tasks. The value will be zero until the Import Review task is completed and the data are successfully loaded.
- **Type** – The import's Media Type is displayed in this column. The possible values are: Data Files, Image Files, Paper Documents, or blank (migrated data). Note: The Image Files value for Media Type will be removed in SEER*DMS version 4.0. This designation was intended for data entry from data submitted to the registry via PDF or other electronic format. It is not used for autoloaded image files which attach scanned document files to SEER*DMS records and patient sets.
- **Comment** – The import-wide comment entered on the same page as contact when starting a new import or entered in the <> tag in the autoloader XML.
- **Import Summary** – This summary shows the total number of records in all data files and the number that were imported (sum of "Valid" for all data files). The difference equals the sum of records with errors, duplicates, and ignored records. Click the **Imported Records** link to view a sample of the records via the Patient Lookup (limited to 500 records).
- **Worklist Summary** – A count for each type of worklist task involving records loaded in this import. You may click the task type to view or open the tasks via the worklist.
- **File Info** – This section shows an error summary for each file involved in this import. See the *Import Summary Information* section of this chapter for a description of the statistics displayed in this section. Click the filename to open or download the original data file.

Finding Records and Tasks Related to an Import

When source data submitted to the registry are imported, the data are assigned unique IDs (with a "REC-" prefix) and are stored in the record table of the SEER*DMS database. In some registries, a pre-load matching algorithm is used for supplemental data types and the non-matching supplemental records are not stored in the database. Refer to your registry-specific Technical Reference and workflow diagrams to determine if all imported supplemental data are stored.

Each imported record travels through the workflow, triggering the automated and manual "tasks" that must be performed to process the data. There are a variety of tools in SEER*DMS for determining the current location of a record in the workflow and the end result of the workflow processing. Use these tools to determine whether there are open tasks related to the records and if the records have been consolidated into the patient set data.

To view a summary of open tasks related to the import:

1. Select **System > Import**. The most recent imports will be listed on the first page of the manager. The pages of the managers include all data entry sessions, autoloaded data, and data loaded in the system interface. You may use the filters to search for a specific

import. The filters correspond to the fields described in the *Data Shown in the Import Manager* section of this chapter.

2. Click the **Import's ID**.
3. The **Worklist Summar**y on the right side of the screen will be recalculated.
4. Click one of the links to view or access the tasks via the worklist.

To search the worklist for open tasks related to the import:

1. Select **View > Worklist**, or click the **Worklist** link on the Home page.
2. Use the **Import** filter to display tasks related to records loaded in a specific import. Enter the Import ID or click the Lookup icon to search for the Import ID by facility, date, and/or user.
3. Some filters are pre-set when you enter the worklist. To ensure that all tasks related to records in this import are listed:

   a. Do not select a value in the task Type(s) filter.

   b. Set the **User(s)** filter to blank.

   c. Check **Show Unassigned**.
4. Click **Apply**.
5. The results will only include record-based tasks for records loaded in the import. When a record moves through the workflow and is linked to a patient set, the focus of the task switches from the record to the patient set. The task would no longer be shown in a list of tasks related to the Import ID. Use RPT-061A to obtain a complete list of tasks initiated by records in an import.

To review a sample of records loaded in a specific import:

1. You may initiate your search from the Import Manager or the Patient Lookup:

   a. To search from the Import Manager:

      i. Select **System > Import Manager**.
      ii. Click the **Lookup** link listed for the import of interest.
      iii. A Patient Lookup search will be executed using the import's ID as a search field. All other search fields are ignored when an Identifier Search field is used.

   b. To initiate the search from the Patient Lookup:

      i. Select **View > Patients**.
      ii. Use the lookup to specify an **Import ID** in the Identifier Search section of the page. All other search fields will be ignored.
      iii. Click **Search**
2. The search results will be limited to a sample of 500 records from the import.

Reports Related to Data Imports

Select **View > Reports** to review the current list of system reports provided in SEER*DMS. The reports listed below are the system reports related to data imports that were included in SEER*DMS when this manual was printed. If you require information regarding imported data that is not included in a system report, you may generate an external report. See *Chapter 24: Creating Reports and Extracting Data* and *SEER*DMS Data Dictionary* for more information.

Report ID	Title	Description
RPT-009A	Imported Records	A list of records imported into SEER*DMS on a specified date. You may select a specific facility, if appropriate.
RPT-061A	Workflow Location of Records	For each record imported at a specified time, this report lists the current location of the record in the workflow.
RPT-070	Import Summary	Number of records loaded per imported data file.

Defining New Data Files Types

Data files imported into SEER*DMS, either within the application or via the auto-loader, must be defined to the import engine in XML configuration files. The XML definitions specify the file layout and conversion rules for the data files. All supported file types are defined in the system configuration. Instructions for viewing these files are provided below. You may define new file types or modify the definitions of existing file types by editing these external configuration files. Alternatively, files may be pre-processed to meet the specifications of a defined file type. External applications, written in any programming language, may be used to perform the necessary reformatting steps.

To view the XML that defines an import type:

1. Select **System > Administration**.
2. Select **System Files** in the **Name** drop-down list.
3. Using the **Files** drop-down list, select one of the XML files listed under **Importer**. Within each file, look at the *name* and *id* attributes to determine the file type defined in the XML.

For more information, refer to the *SEER*DMS Technical Reference*.